PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C07K 14/00, 16/00, C12N 5/00, 15/00, C07H 21/00, A61K 31/00, 38/00, 48/00, C12Q 1/68, C12P 19/34, G01N 33/00

A1

(11) International Publication Number: WO 96/26215

(43) International Publication Date: 29 August 1996 (29.08.96)

(21) International Application Number: PCT/US96/01528

(22) International Filing Date: 12 February 1996 (12.02.96)

(30) Priority Data:

| | | |
|---|---|---|
| 08/401,530 | 22 February 1995 (22.02.95) | US |
| 60/006,271 | 7 November 1995 (07.11.95) | US |

(81) Designated States: AL, AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TT, UA, UG, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AZ, BY, KG, KZ, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
*With international search report.*

(71) Applicants: EASTERN VIRGINIA MEDICAL SCHOOL OF THE MEDICAL COLLEGE OF HAMPTON ROADS [US/US]; P.O. Box 1980, Norfolk, VA 23501 (US). McGill University [CA/CA]; 845 Sherbrooke Street West, Montreal, Quebec H3A 2T5 (CA).

(72) Inventors: VINIK, Aaron, I.; 40 Rader Street #603, Norfolk, VI 23510 (US). PITTENGER, Gary, L.; 3701 Prince Andrew Lane, Virginia Beach, VI 23452 (US). RAFAELOFF, Ronit; 1040 Spotswood Avenue #102, Norfolk, VI 23507 (US). ROSENBERG, Lawrence; 6507 Fern Road, Montreal, Quebec H4V 1E4 (CA). DUGUID, William, P.; 209 Calais Drive, Baie d'Urfe, Montreal, Quebec H9X 2L6 (CA).

(74) Agents: KAGAN, Sarah, A. et al.; Banner & Allegretti, Ltd., 11th floor, 1001 G Street, N.W., Washington, DC 20001 (US).

(54) Title: INGAP PROTEIN INVOLVED IN PANCREATIC ISLET NEOGENESIS

(57) Abstract

Cellophane wrapping (CW) of hamster pancreas induces proliferation of duct epithelial cells followed by endocrine cell differentiation and islet neogenesis. Using the mRNA differential display technique a cDNA clone expressed in cellophane wrap but not in control pancreata was identified. Using this cDNA as a probe, a cDNA library was screened and a gene not previously described was identified and named *INGAP*.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | |
|---|---|---|---|---|---|
| **AM** | Armenia | **GB** | United Kingdom | **MW** | Malawi |
| **AT** | Austria | **GE** | Georgia | **MX** | Mexico |
| **AU** | Australia | **GN** | Guinea | **NE** | Niger |
| **BB** | Barbados | **GR** | Greece | **NL** | Netherlands |
| **BE** | Belgium | **HU** | Hungary | **NO** | Norway |
| **BF** | Burkina Faso | **IE** | Ireland | **NZ** | New Zealand |
| **BG** | Bulgaria | **IT** | Italy | **PL** | Poland |
| **BJ** | Benin | **JP** | Japan | **PT** | Portugal |
| **BR** | Brazil | **KE** | Kenya | **RO** | Romania |
| **BY** | Belarus | **KG** | Kyrgystan | **RU** | Russian Federation |
| **CA** | Canada | **KP** | Democratic People's Republic of Korea | **SD** | Sudan |
| **CF** | Central African Republic | | | **SE** | Sweden |
| **CG** | Congo | **KR** | Republic of Korea | **SG** | Singapore |
| **CH** | Switzerland | **KZ** | Kazakhstan | **SI** | Slovenia |
| **CI** | Côte d'Ivoire | **LI** | Liechtenstein | **SK** | Slovakia |
| **CM** | Cameroon | **LK** | Sri Lanka | **SN** | Senegal |
| **CN** | China | **LR** | Liberia | **SZ** | Swaziland |
| **CS** | Czechoslovakia | **LT** | Lithuania | **TD** | Chad |
| **CZ** | Czech Republic | **LU** | Luxembourg | **TG** | Togo |
| **DE** | Germany | **LV** | Latvia | **TJ** | Tajikistan |
| **DK** | Denmark | **MC** | Monaco | **TT** | Trinidad and Tobago |
| **EE** | Estonia | **MD** | Republic of Moldova | **UA** | Ukraine |
| **ES** | Spain | **MG** | Madagascar | **UG** | Uganda |
| **FI** | Finland | **ML** | Mali | **US** | United States of America |
| **FR** | France | **MN** | Mongolia | **UZ** | Uzbekistan |
| **GA** | Gabon | **MR** | Mauritania | **VN** | Viet Nam |

# INGAP PROTEIN INVOLVED IN PANCREATIC ISLET NEOGENESIS

## BACKGROUND OF THE INVENTION

Pancreatic islets of *Langerhans* are the only organ of insulin production in the body. However, they have a limited capacity for regeneration. This limited regeneration capacity predisposes mammals to develop diabetes mellitus. Thus there is a need in the art of endocrinology for products which can stimulate the regeneration of islets of *Langerhans* to prevent or ameliorate the symptoms of diabetes mellitus.

One model of pancreatic islet cell regeneration involves cellophane-wrapping of the pancreas in the Syrian golden hamster (1). Wrapping of the pancreas induces the formation of new endocrine cells which appear to arise from duct epithelium (2-4). There is a need in the art to identify and isolate the factor(s) which is responsible for islet cell regeneration.

## SUMMARY OF THE INVENTION

It is an object of the invention to provide a preparation of a mammalian protein or polypeptide portions thereof involved in islet cell neogenesis.

It is another object of the invention to provide a DNA molecule encoding a mammalian protein involved in islet cell neogenesis.

It is yet another object of the invention to provide a preparation of a mammalian INGAP (islet neogenesis associated protein) protein.

It is still another object of the invention to provide nucleotide probes for detecting mammalian genes involved in islet cell neogenesis.

It is an object of the invention to provide a method for isolation of INGAP genes from a mammal.

It is another object of the invention to provide an antibody preparation which is specifically immunoreactive with an INGAP protein.

It is yet another object of the invention to provide methods of producing INGAP proteins.

It is an object of the invention to provide methods for treating diabetic mammals.

It is another object of the invention to provide methods for growing pancreatic islet cells in culture.

It is still another object of the invention to provide methods of enhancing the life span of pancreatic islet cells encapsulated in polycarbon shells.

It is an object of the invention to provide methods of enhancing the number of pancreatic islet cells in a mammal.

It is an object of the invention to provide transgenic mammals.

It is another object of the invention to provide genetically engineered mammals.

It is yet another object of the invention to provide methods of identifying individual mammals at risk for diabetes.

It is an object of the invention to provide methods of detecting INGAP protein in a sample from a mammal.

It is still another object of the invention to provide a method of treating isolated islet cells to avoid apoptosis.

It is another object of the invention to provide methods of treating mammals receiving islet cell transplants.

It is an object of the invention to provide a method of inducing differentiation of $\beta$ cell progenitors.

It is an object of the invention to provide a method of identifying $\beta$ cell progenitors.

It is another object of the invention to provide a method of treating a mammal with pancreatic endocrine failure.

It is an object of the invention to provide antisense constructs for regulating the expression of *INGAP*.

It is yet another object of the invention to provide a method for treating nesidioblastosis.

It is still another object of the invention to provide kits for detecting mammalian INGAP proteins.

It is an object of the invention to provide pharmaceutical compositions for treatment of pancreatic insufficiency.

These and other objects of the invention are provided by one or more of the embodiments described below.

In one embodiment a preparation of a mammalian INGAP protein is provided. The preparation is substantially free of other mammalian proteins.

In another embodiment an isolated cDNA molecule is provided. The cDNA molecule encodes a mammalian INGAP protein.

In still another embodiment of the invention a preparation of a mammalian INGAP protein is provided. The preparation is made by the process of:

inducing mammalian pancreatic cells to express INGAP protein by cellophane-wrapping; and

purifying said INGAP protein from said induced mammalian pancreatic cells.

In yet another embodiment of the invention a nucleotide probe is provided. The probe comprises at least 20 contiguous nucleotides of the sequence shown in SEQ ID NO: 1.

In another embodiment of the invention a preparation of INGAP protein of a mammal is provided. The preparation is substantially purified from other proteins of the mammal. The INGAP protein is inducible upon cellophane-wrapping of pancreas of the mammal.

In yet another embodiment of the invention a method of isolating an *INGAP* gene from a mammal is provided. The method comprises:

hybridizing one or more oligonucleotides comprising at least 10 contiguous nucleotides of the sequence shown in SEQ ID NO: 1 to genomic DNA or cDNA of said mammal;

identifying DNA molecules from said genomic DNA or cDNA which hybridize to said one or more oligonucleotides.

In still another embodiment of the invention an isolated cDNA molecule is provided. The cDNA molecule is obtained by the process of:

hybridizing one or more oligonucleotides comprising at least 10 contiguous nucleotides of the sequence shown in SEQ ID NO: 1 to genomic DNA or cDNA of said mammal;

identifying DNA molecules from said genomic DNA or cDNA which hybridize to said one or more oligonucleotides.

In another embodiment of the invention an antibody is provided. The antibody is specifically immunoreactive with a mammalian INGAP protein.

According to still another embodiment of the invention a method of producing a mammalian INGAP protein is provided. The method comprises the steps of:

providing a host cell transformed with a cDNA encoding a mammalian INGAP protein;

culturing the host cell in a nutrient medium so that the INGAP protein is expressed; and

harvesting the INGAP protein from the host cell or the nutrient medium.

According to yet another embodiment of the invention a method of producing a mammalian INGAP protein is provided. The method comprises the steps of:

providing a host cell comprising a DNA molecule obtained by the process of:

hybridizing one or more oligonucleotides comprising at least 10 contiguous nucleotides of the sequence shown in SEQ ID NO: 1 to genomic DNA or cDNA of said mammal;

identifying DNA molecules from said genomic DNA or cDNA which hybridize to said one or more oligonucleotides;

culturing the host cell in a nutrient medium so that the mammalian INGAP protein is expressed; and

harvesting the mammalian INGAP protein from the host cells or the nutrient medium.

According to another embodiment of the invention a method of treating diabetic mammals is provided. The method comprises:

administering to a diabetic mammal a therapeutically effective amount of an INGAP protein to stimulate growth of islet cells.

According to another embodiment of the invention a method of growing pancreatic islet cells in culture is provided. The method comprises:

supplying an INGAP protein to a culture medium for growing pancreatic islet cells; and

growing islet cells in said culture medium comprising INGAP protein.

According to another embodiment of the invention a method of enhancing the life span of pancreatic islet cells encapsulated in a polycarbon shell is provided. The method comprises:

adding to encapsulated pancreatic islet cells an INGAP protein in an amount sufficient to enhance the survival rate or survival time of said pancreatic islet cells.

According to another embodiment of the invention a method of enhancing the number of pancreatic islet cells in a mammal is provided. The method comprises:

administering a DNA molecule which encodes an INGAP protein to a pancreas in a mammal.

According to another embodiment of the invention a method of enhancing the number of pancreatic islet cells in a mammal is provided. The method comprises:

administering an INGAP protein to a pancreas in a mammal.

According to another embodiment of the invention a transgenic mammal is provided. The mammal comprises an *INGAP* gene of a second mammal.

According to another embodiment of the invention a non-human mammal is provided. The mammal has been genetically engineered to contain an insertion or deletion mutation of an *INGAP* gene of said mammal.

According to another embodiment of the invention a method of identifying individual mammals at risk for diabetes is provided. The method comprises:

identifying a mutation in an *INGAP* gene of a sample of an individual mammal, said mutation causing a structural abnormality in an INGAP protein encoded by said gene or causing a regulatory defect leading to diminished or obliterated expression of said *INGAP* gene.

According to another embodiment of the invention a method of detecting INGAP protein in a sample from a mammal is provided. The method comprises:

contacting said sample with an antibody preparation which is specifically immunoreactive with a mammalian INGAP protein.

According to another embodiment of the invention a method of treating isolated islet cells of a mammal to avoid apoptosis of said cells is provided. The method comprises:

contacting isolated islet cells of a mammal with a preparation of a mammalian INGAP protein, substantially purified from other mammalian proteins, in an amount sufficient to increase the survival rate of said isolated islet cells.

According to another embodiment of the invention a method of treating a mammal receiving a transplant of islet cells is provided. The method comprises:

administering a preparation of a mammalian INGAP protein to a mammal receiving a transplant of islet cells, wherein said step of administering is performed before, during, or after said transplant.

According to another embodiment of the invention a method of inducing differentiation of $\beta$ cell progenitors is provided. The method comprises:

contacting a culture of pancreatic duct cells comprising β cell progenitors with a preparation of a mammalian INGAP protein substantially free of other mammalian proteins, to induce differentiation of said β cell progenitors.

In yet another embodiment of the invention a method is provided for identification of β cell progenitors. The method comprises:

contacting a population of pancreatic duct cells with a mammalian INGAP protein; and

detecting cells among said population to which said INGAP protein specifically binds.

According to another embodiment of the invention a method of treating a mammal with pancreatic endocrine failure is provided. The method comprises:

contacting a preparation of pancreatic duct cells comprising β cell progenitors isolated from a mammal afflicted with pancreatic endocrine failure with a preparation of a mammalian INGAP protein substantially free of other mammalian proteins to induce differentiation of said β cell progenitors; and

autologously transplanting said treated pancreatic duct cells into said mammal.

According to another embodiment of the invention an antisense construct of a mammalian *INGAP* gene is provided. The construct comprises:

a promoter, a terminator, and a nucleotide sequence consisting of a mammalian *INGAP* gene, said nucleotide sequence being between said promoter and said terminator, said nucleotide sequence being inverted with respect to said promoter, whereby upon expression from said promoter an mRNA complementary to native mammalian *INGAP* mRNA is produced.

According to another embodiment of the invention a method of treating nesidioblastosis is provided. The method comprises:

administering to a mammal with nesidioblastosis an antisense construct as described above, whereby overgrowth of β cells of said mammal is inhibited.

According to another embodiment of the invention a kit for detecting a mammalian INGAP protein in a sample from a mammal is provided. The kit comprises:

an antibody preparation which is specifically immunoreactive with a mammalian INGAP protein; and

a polypeptide which comprises a sequence of at least 15 consecutive amino acids of a mammalian INGAP protein.

According to another embodiment of the invention a pharmaceutical composition for treatment of pancreatic insufficiency is provided. The composition comprises:

a mammalian INGAP protein in a pharmaceutically acceptable diluent or carrier.

According to another embodiment of the invention a pharmaceutical composition is provided. The composition comprises:

a preparation of a polypeptide which comprises a sequence of at least 15 consecutive amino acids of a mammalian INGAP protein and a pharmaceutically acceptable diluent or carrier.

These and other embodiments of the invention provide the art with means of stimulating and inhibiting islet cell neogenesis. Means of diagnosis of subsets of diabetes mellitus are also provided by this invention.

## BRIEF DESCRIPTION OF THE DRAWINGS

**Figure 1.** *Nucleotide sequence of hamster INGAP* and deduced sequence of encoded immature protein. The non-coding sequences are in lower case letters, and the polyadenylation signal is underlined.

**Figure 2.** *Comparison of amino acid sequences of INGAP*, rat PAP-I (PAP-I) (18), Human PAP/HIP (PAP-H/HIP)(10,11), rat PAP-III (PAP-III)(9), rat PAP-II (PAP-II)(8), Rat Reg/PSP/Lithostatine (REG/LITH)(13,15) and the

invariable motif found by Drickamer in all members of C-type lectins (Drickamer) (12). Six conserved cysteines are marked by asterisks and the 2 putative *N*-glycosylation sites of INGAP are underlined and in bold letters.

**Figure 3.** ***Northern blot analysis of INGAP*** and amylase gene expression in pancreatic tissue from control and wrapped hamster pancreas. 30 g of heat denatured total RNA was separated by electrophoresis on a 1.2% agarose, 0.6% formaldehyde/MOPS denaturing gel, and transferred to nylon membrane. Membranes were hybridized with a 747bp hamster *INGAP* cDNA probe (cloned in our lab) (A), a 1000bp rat amylase cDNA probe (generously given by Chris Newgard Dallas, Texas) (B) and with an 18S ribosomal 24mer synthetic oligonucleotide probe to control for RNA integrity and loading (C).

## DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

We now report the identification of a gene, *INGAP*, that shows striking homology to the pancreatitis associated protein (PAP) family of genes (7-11). The predicted protein shares the carbohydrate recognition domain (CRD) of the calcium dependent C-type lectins as defined by Drickamer (12). INGAP protein plays a role in stimulation of islet neogenesis, in particular, in beta cell regeneration from ductal cells.

The cDNA sequence of a mammalian *INGAP* is provided in SEQ ID NO: 1. The predicted amino acid sequence is shown in SEQ ID NO:2. These sequences were determined from nucleic acids isolated from hamster, but it is believed that other mammalian species will contain *INGAP* genes which are quite similar. Human *INGAP* cDNA shares the sequence from 23 to 268, and from 389 to 609 in SEQ ID NO:1 with a 159 bp gap in the middle of the sequence. The predicted amino acid sequence of human *INGAP* protein is from 1 to 83, and from 124 to 174 in SEQ ID NO:2 with 53 amino acids in the middle of the sequence. One would expect homologous genes to contain at least about 70% identity. Closer species would be expected to have at least about 75%, 80%, or even 85%

identity. In contrast, other family members of the calcium dependent C-type lectins contain at most 60% identity with *INGAP*.

The DNA sequence provided herein can be used to form vectors which will replicate the gene in a host cell, and may also express INGAP protein. DNA sequences which encode the same amino acid sequence as shown in SEQ ID NO:2 can also be used, without departing from the contemplation of the invention. DNA sequences coding for other mammalian INGAPs are also within the contemplation of the invention. Suitable vectors, for both prokaryotic and eukaryotic cells, are known in the art. Some vectors are specifically designed to effect expression of inserted DNA segments downstream from a transcriptional and translational control site. One such vector for expression in eukaryotic cells employs EBNA His, a plasmid which is available commercially from InVitrogen Corp. The loaded vector produces a fusion protein comprising a portion of a histidine biosynthetic enzyme and INGAP. Another vector, which is suitable for use in prokaryotic cells, is pCDNA3. Selection of a vector for a particular purpose may be made using knowledge of the properties and features of the vectors, such as useful expression control sequences. Vectors may be used to transform or transfect host cells, either stably or transiently. Methods of transformation and transfection are known in the art, and may be used according to suitability for a particular host cell. Host cells may be selected according to the purpose of the transfection. A suitable prokaryotic host is *E. coli* DH5$\alpha$. A suitable eukaryotic host is cos7, an African Green Monkey kidney cell line. For some purposes, proper glycosylation of INGAP may be desired, in which case a suitable host cell should be used which recognizes the glycosylation signal of *INGAP*.

Probes comprising at least 10, 15, 20, or 30 nucleotides of contiguous sequence according to SEQ ID NO:1 can be used for identifying *INGAP* genes in particular individuals or in members of other species. Appropriate conditions for hybridizations to same or different species' DNA are known in the art as high stringency and low stringency, respectively. These can be used in a variety of

formats according to the desired use. For example, Southern blots, Northern blots, and *in situ* colony hybridization, can be used as these are known in the art. Probes typically are DNA or RNA oligomers of at least 10, 15, 20, or 30 nucleotides. The probe may be labeled with any detectable moiety known in the art, including radiolabels, fluorescent labels, enzymes, etc. Probes may also be derived from other mammalian *INGAP* gene sequences.

*INGAP* genes can be isolated from other mammals by utilizing the nucleotide sequence information provided herein. (More laboriously, they can be isolated using the same method described in detail below for isolation of the hamster *INGAP* gene.) Oligonucleotides comprising at least 10 contiguous nucleotides of the disclosed nucleotide sequence of *INGAP* are hybridized to genomic DNA or cDNA of the mammal. The DNA may conveniently be in the form of a library of clones. The oligonucleotides may be labelled with any convenient label, such as a radiolabel or an enzymatic or fluorescence label. DNA molecules which hybridize to the probe are isolated. Complete genes can be constructed by isolating overlapping DNA segments, for example using the first isolated DNA as a probe to contiguous DNA in the library or preparation of the mammal's DNA. Confirmation of the identity of the isolated DNA can be made by observation of the pattern of expression of the gene in the pancreas when subjected to cellophane wrapping, for example. Similarly, the biological effect of the encoded product upon pancreatic ductal cells will also serve to identify the gene as an *INGAP* gene.

If two oligonucleotides are hybridized to the genomic DNA or cDNA of the mammal then they can be used as primers for DNA synthesis, for example using the polymerase chain reaction or the ligase chain reaction. Construction of a full-length gene and confirmation of the identity of the isolated gene can be performed as described above.

INGAP protein may be isolated according to the invention by inducing mammalian pancreatic cells to express INGAP protein by means of cellophane-

wrapping. This technique is described in detail in reference no. 1 which is expressly incorporated herein. INGAP protein so produced may be purified from other mammalian proteins by means of immunoaffinity techniques, for example, or other techniques known in the art of protein purification. An antibody specific for a mammalian INGAP is produced using all, or fragments of, the amino acid sequence of an INGAP protein, such as shown in SEQ ID NO: 2, as immunogens. The immunogens can be used to identify and purify immunoreactive antibodies. Monoclonal or polyclonal antibodies can be made as is well known in the art. The antibodies can be conjugated to other moieties, such as detectable labels or solid support materials. Such antibodies can be used to purify proteins isolated from mammalian pancreatic cells or from recombinant cells. Hybridomas which secrete specific antibodies for an INGAP protein are also within the contemplation of the invention.

Host cells as described above can be used to produce a mammalian INGAP protein. The host cells comprise a DNA molecule encoding a mammalian INGAP protein. The DNA can be according to SEQ ID NO:1, or isolated from other mammals according to methods described above. Host cells can be cultured in a nutrient medium under conditions where INGAP protein is expressed. INGAP protein can be isolated from the host cells or the nutrient medium, if the INGAP protein is secreted from the host cells.

It has now been found that INGAP and fragments thereof are capable of inducing and stimulating islet cells to grow. Moreover, they are capable of inducing differentiation of pancreatic duct cells, and of allowing such cells to avoid the apoptotic pathway. Thus many therapeutic modalities are now possible using INGAP, fragments thereof, and nucleotide sequences encoding INGAP. Therapeutically effective amounts of INGAP are supplied to patient pancreata, to isolated islet cells, and to encapsulated pancreatic islet cells, such as in a polycarbon shell. Suitable amounts of INGAP for therapeutic purposes range from 1-150 $\mu$g/kg of body weight or *in vitro* from 1-10,000 $\mu$g/ml. Optimization of

such dosages can be ascertained by routine testing. Methods of administering INGAP to mammals can be any that are known in the art, including subcutaneous, via the portal vein, by local perfusion, etc.

Conditions which can be treated according to the invention by supplying INGAP include diabetes mellitus, both insulin dependent and non-insulin dependent, pancreatic insufficiency, pancreatic failure, etc. Inhibition of INGAP expression can be used to treat nesidioblastosis.

According to the present invention, it has now been found that a small portion of INGAP is sufficient to confer biological activity. A fragment of 20 amino acids of the sequence of SEQ ID NO: 2, from amino acid #103-#122 is sufficient to stimulate pancreatic ductal cells to grow and proliferate. The effect has been seen on a rat tumor duct cell line, a hamster duct cell line, a hamster insulinoma cell line, and a rat insulinoma cell line. The analogous portions of other mammalian INGAP proteins are quite likely to have the same activity. This portion of the protein is not similar to other members of the pancreatitis associated protein (PAP) family of proteins. It contains a glycosylation site and it is likely to be a primary antigenic site of the protein as well. This fragment has been used to immunize mice to generate monoclonal antibodies.

The physiological site of expression of INGAP has been determined. INGAP is expressed in acinar tissue, in the exocrine portion of the pancreas. It is not expressed in ductal or islet cells, *i.e.*, the paracrine portion of the pancreas. Expression occurs within 24-48 hours of induction by means of cellophane wrapping.

Transgenic animals according to the present invention are mammals which carry an *INGAP* gene from a different mammal. The transgene can be expressed to a higher level than the endogenous *INGAP* genes by judicious choice of transcription regulatory regions. Methods for making transgenic animals are well-known in the art, and any such method can be used. Animals which have been genetically engineered to carry insertions, deletions, or other mutations which alter

the structure of the INGAP protein or regulation of expression of *INGAP* are also contemplated by this invention. The techniques for effecting these mutations are known in the art.

Diagnostic assays are also contemplated within the scope of the present invention. Mutations in *INGAP* can be ascertained in samples such as blood, amniotic fluid, chorionic villus, blastocyst, and pancreatic cells. Such mutations identify individuals who are at risk for diabetes. Mutations can be identified by comparing the nucleotide sequence to a wild-type sequence of an *INGAP* gene. This can be accomplished by any technique known in the art, including comparing restriction fragment length polymorphisms, comparing polymerase chain reaction products, nuclease protection assays, etc. Alternatively, altered proteins can be identified, *e.g.*, immunologically or biologically.

The present invention also contemplates the use of *INGAP* antisense constructs for treating nesidioblastosis, a condition characterized by overgrowth of $\beta$ cells. The antisense construct is administered to a mammal having nesidioblastosis, thereby inhibiting the overgrowth of $\beta$ cells. An antisense construct typically comprises a promoter, a terminator, and a nucleotide sequence consisting of a mammalian *INGAP* gene. The *INGAP* sequence is between the promoter and the terminator and is inverted with respect to the promoter as it is expressed naturally. Upon expression from the promoter, an mRNA complementary to native mammalian *INGAP* is produced.

Immunological methods for assaying INGAP in a sample from a mammal are useful, for example, to monitor the therapeutic administration of INGAP. Typically an antibody specific for INGAP will be contacted with the sample and the binding between the antibody and any INGAP in the sample will be detected. This can be by means of a competitive binding assay, in which the incubation mixture is spiked with a known amount of a standard INGAP preparation, which may conveniently be detectably labeled. Alternatively, a polypeptide fragment of INGAP may be used as a competitor. In one particular assay format, the

antibodies are bound to a solid phase or support, such as a bead, polymer matrix, or a microtiter plate.

According to the present invention, pancreatic duct cells of a mammal with pancreatic endocrine failure can be removed from the body and treated *in vitro*. The duct cells typically comprise $\beta$ cell progenitors. Thus treatment with a preparation of a mammalian INGAP protein will induce differentiation of the $\beta$ cell progenitors. The duct cells are contacted with a preparation of a mammalian INGAP protein substantially free of other mammalian proteins. The treated cells can then used as an autologous transplant into the mammal from whom they were derived. Such an autologous treatment minimizes adverse host versus graft reactions involved in transplants.

INGAP protein can also be used to identify those cells which bear receptors for INGAP. Such cells are likely to be the $\beta$ cell progenitors, which are sensitive to the biological effects of INGAP. INGAP protein can be detectably labeled, such as with a radiolabel or a fluorescent label, and then contacted with a population of cells from the pancreatic duct. Cells which bind to the labeled protein will be identified as those which bear receptors for INGAP, and thus are $\beta$ cell progenitors. Fragments of INGAP can also be used for this purpose, as can immobilized INGAP which can be used to separate cells from a mixed population of cells to a solid support. INGAP can be immobilized to solid phase or support by adsorption to a surface, by means of an antibody, or by conjugation. Any other means as is known in the art can also be used.

Kits are provided by the present invention for detecting a mammalian INGAP protein in a sample. This may be useful, *inter alia*, for monitoring metabolism of INGAP during therapy which involves administration of INGAP to a mammal. The kit will typically contain an antibody preparation which is specifically immunoreactive with a mammalian INGAP protein. The antibodies may be polyclonal or monoclonal. If polyclonal they may be affinity purified to render them monospecific. The kit will also typically contain a polypeptide which

has at least 15 consecutive amino acids of a mammalian INGAP protein. The polypeptide is used to compete with the INGAP protein in a sample for binding to the antibody. Desirably the polypeptide will be detectably labeled. The polypeptide will contain the portion of INGAP to which the antibody binds. Thus if the antibody is monoclonal, the polypeptide will successfully compete with INGAP by virtue of it containing the epitope of the antibody. It may also be desirable that the antibodies be bound to a solid phase or support, such as polymeric beads, sticks, plates, etc.

Pharmaceutical compositions containing a mammalian INGAP protein may be used for treatment of pancreatic insufficiency. The composition may alternatively contain a polypeptide which contains a sequence of at least 15 consecutive amino acids of a mammalian INGAP protein. The polypeptide will contain a portion of INGAP which is biologically active in the absence of the other portions of the protein. The polypeptide may be part of a larger protein, such as a genetic fusion with a second protein or polypeptide. Alternatively, the polypeptide may be conjugated to a second protein, for example, by means of a cross-linking agent. Suitable portions of INGAP proteins may be determined by homology with amino acids #103 to #122 of SEQ ID NO:2, or by the ability of test polypeptides to stimulate pancreatic duct cells to grow and proliferate. As is known in the art, it is often the case that a relatively small number of amino acids can be removed from either end of a protein without destroying activity. Thus it is contemplated within the scope of the invention that up to about 10% of the protein can be deleted, and still provide essentially all functions of INGAP. Such proteins have at least about 130 amino acids, in the case of hamster INGAP.

The pharmaceutical composition will contain a pharmaceutically acceptable diluent or carrier. A liquid formulation is generally preferred. INGAP may be formulated at different concentrations or using different formulants. For example, these formulants may include oils, polymers, vitamins, carbohydrates, amino acids, salts, buffers, albumin, surfactants, or bulking agents. Preferably

carbohydrates include sugar or sugar alcohols such as mono-, di-, or polysaccharides, or water soluble glucans. The saccharides or glucans can include fructose, dextrose, lactose, glucose, mannose, sorbose, xylose, maltose, sucrose, dextran, pullulan, dextrin, alpha and beta cyclodextrin, soluble starch, hydroxethyl starch and carboxymethylcelloluose, or mixtures thereof. Sucrose is most preferred. Sugar alcohol is defined as a $C_4$ to $C_8$ hydrocarbon having an -OH group and includes galactitol, inositol, mannitol, xylitol, sorbitol, glycerol, and arabitol. Mannitol is most preferred. These sugars or sugar alcohols mentioned above may be used individually or in combination. There is no fixed limit to amount used as long as the sugar or sugar alcohol is soluble in the aqueous preparation. Preferably, the sugar or sugar alcohol concentration is between 1.0 w/v% and 7.0 w/v%, more preferable between 2.0 and 6.0 w/v%. Preferably amino acids include levorotary (L) forms of carnitine, arginine, and betaine; however, other amino acids may be added. Preferred polymers include polyvinylpyrrolidone (PVP) with an average molecular weight between 2,000 and 3,000, or polyethylene glycol (PEG) with an average molecular weight between 3,000 and 5,000. It is also preferred to use a buffer in the composition to minimize pH changes in the solution before lyophilization or after reconstitution, if these are used. Most any physiological buffer may be used, but citrate, phosphate, succinate, and glutamate buffers or mixtures thereof are preferred. Preferably, the concentration is from 0.01 to 0.3 molar. Surfactants can also be added to the formulation.

Additionally, INGAP or polypeptide portions thereof can be chemically modified by covalent conjugation to a polymer to increase its circulating half-life, for example. Preferred polymers, and methods to attach them to peptides, are shown in U.S. Patent Nos. 4,766,106, 4,179,337, 4,495,285, and 4,609,546. Preferred polymers are polyoxyethylated polyols and polyethylene glycol (PEG). PEG is soluble in water at room temperature and has the general formula: $R(O-CH_2-CH_2)_nO-R$ where R can be hydrogen, or a protective group such as an alkyl

or alkanol group. Preferably, the protective group has between 1 and 8 carbons, more preferably it is methyl. The symbol n is a positive integer, preferably between 1 and 1,000, more preferably between 2 and 500. The PEG has a preferred average molecular weight between 1000 and 40,000, more preferably between 2000 and 20,000, most preferably between 3,000 and 12,000. Preferably, PEG has at least one hydroxy group, more preferably it is a terminal hydroxy group. It is this hydroxy group which is preferably activated to react with a free amino group on the inhibitor.

After the liquid pharmaceutical composition is prepared, it is preferably lyophilized to prevent degradation and to preserve sterility. Methods for lyophilizing liquid compositions are known to those of ordinary skill in the art. Just prior to use, the composition may be reconstituted with a sterile diluent (Ringer's solution, distilled water, or sterile saline, for example) which may include additional ingredients. Upon reconstitution, the composition is preferably administered to subjects using those methods that are known to those skilled in the art.

The following examples are not intended to limit the scope of the invention, but merely to exemplify that which is taught above.

## Examples

### Example 1

This example describes the cloning and isolation of a cDNA encoding a novel, developmentally regulated, pancreatic protein.

We hypothesized that a unique locally produced factor(s) is responsible for islet cell regeneration. Using the recently developed mRNA differential display technique (5,6) to compare genes differentially expressed in cellophane wrapped (CW) versus control pancreata (CP) allowed us to identify a cDNA clone (RD19-2) which was uniquely expressed in cellophane wrapped pancreas.

A cDNA library was constructed from mRNA isolated from cellophane wrapped hamster pancreas using oligo d(T) primed synthesis, and ligation into pcDNA3 vector (Invitrogen). The number of primary recombinants in the library was 1.2 X $10^6$ with an average size of 1.1 kb. The cDNA library was screened for clones of interest using high density colony plating techniques. Colonies were lifted onto nylon membranes (Schleicher & Schuell) and further digested with proteinase K (50(g/ml). Treated membranes were baked at 80°C for 1 hour and hybridized at 50°C for 16-18 hours with 1-5 x $10^6$ cpm/ml of [($^{32}$P]-dCTP(Dupont-New England Nuclear) radiolabeled RD19-2 probe. Colonies with a positive hybridization signal were isolated, compared for size with Northern mRNA transcript, and sequenced to confirm identity with the RD19-2 sequence.

Example 2

This example compares the sequence of INGAP to other proteins with which it shares homology.

The nucleotide sequence of the hamster *INGAP* clone with the longest cDNA insert was determined. As shown in Figure 1 the hamster cDNA comprises 747 nucleotides (nt), exclusive of the poly(A) tail and contains a major open reading frame encoding a 175 amino acid protein. The open reading frame is followed by a 3'-untranslated region of 206nt. A typical polyadenylation signal is present 11nt upstream of the poly(A) tail. The predicted INGAP protein shows structural homology to both the PAP/HIP family of genes which is associated with pancreatitis or liver adenocarcinoma (7-11) and the Reg/PSP/lithostatine family of genes (13,15) which has been shown to stimulate pancreatic beta-cell growth (14) and might play a role in pancreatic islet regeneration. Comparison of the nucleotide sequence and their deduced amino acids between hamster INGAP and rat PAP-I shows a high degree of homology in the coding region (60 and 58% in nucleotide and amino acid sequences, respectively). The predicted amino acid sequence of the hamster INGAP reveals 45% identity to PAP II and 50% to PAP III both of which have been associated with acute pancreatitis, and 54% to HIP

which was found in a hepato-cellular carcinoma. INGAP also shows 40% identity to the rat Reg/PSP/lithostatine protein (Fig. 2). Reg is thought to be identical to the pancreatic stone protein (PSP) (15,16) or pancreatic thread protein (PTP) (17). The *N*-terminus of the predicted sequence of INGAP protein is highly hydrophobic which makes it a good candidate for being the signal peptide which would allow the protein to be secreted. Similar to PAP/HIP but different from the Reg/PSP/lithostatine proteins a potential *N*-glycosylation site is situated at position 135 of the INGAP sequence. Unique to INGAP is another potential *N*-glycosylation site situated at position 115. INGAP also shows a high degree of homology (12/18) (Fig. 2) with a consensus motif in members of the calcium-dependent (C-type) animal lectin as determined by Drickamer including four perfectly conserved cysteines which form two disulfide bonds (12). Two extra cysteines found at the amino-terminus of INGAP (Fig. 2) are also present in Reg/PSP and PAP/HIP. However, it is not clear what the biological significance might be.

Example 3

This example demonstrates the temporal expression pattern of *INGAP* upon cellophane-wrapping.

In order to determine the temporal expression of the *INGAP* gene, total RNA extracted from CP and CW pancreas was probed with the hamster *INGAP* cDNA clone in Northern blot analysis. A strong single transcript of 900bp was detected (Fig. 3) 1 and 2 days after cellophane wrapping which disappeared by 6 through 42 days and was absent from CP. *INGAP* mRNA is associated with CW induced pancreatic islet neogenesis, since it is present only after CW. It is not likely that the increased expression of *INGAP* is associated with acute pancreatitis as is the case with the PAP family of genes. During the acute phase of pancreatitis the concentrations of most mRNAs encoding pancreatic enzymes including amylase are decreased significantly (16,18). In contrast, in the CW model of islet neogenesis in which high expression of *INGAP* has been detected,

amylase gene expression was simultaneously increased above normal (Fig. 3) rather than decreased, suggesting that *INGAP* expression is not associated with pancreatitis but rather with islet neogenesis. The cause of increased amylase gene expression 1 and 2 days after CW is as yet unclear, and more studies need to be done to elucidate this issue. It is unlikely though, that the increase is associated with exocrine cell regeneration which occurs at a later time after CW (19). Thus, INGAP protein plays a role in stimulation of islet neogenesis, in particular, in beta cell regeneration from ductal cells.

Example 4

This example describes the cloning and partial sequence of a human cDNA encoding INGAP protein.

Human polyA$^+$ RNA was isolated from a normal human pancreas using a commercially available polyA$^+$ extraction kit from Qiagen. Subsequently, 500 ng polyA$^+$ RNA was used as a template for reverse transcription and polymerase chain reaction (RT-PCR). The experimental conditions were set according to the instructions in the RT-PCR kit from Perkin Elmer. Oligo d(T) was used as the primer in reverse transcription. Primers corresponding to nucleotides 4 to 23 and 610 to 629 in SEQ ID NO:1 were used as the specific primers in the polymerase chain reaction. A 626 bp PCR fragment was cloned using a TA cloning kit from Invitrogen. The partial sequence of the human clone comprises 466 bp with a 120 bp gap in the middle of the sequence. The human *INGAP* cDNA is 100% identical to the hamster *INGAP* cDNA sequence from nucleotide 4 to 268, and from nucleotide 289 to 629 in SEQ ID NO:1. The sequence of the 120 bp in the middle is as yet unidentified.

Example 5

This example demonstrates that synthetic peptides from *INGAP* play a role in stimulation of islet neogenesis, and that at least one epitope coded by the as yet unsequenced 120 bp segment of human *INGAP* is shared with hamster *INGAP*.

A synthetic peptide corresponding to amino acids 104-118 in SEQ ID NO:2 of the deduced hamster *INGAP* protein was used as an immunogen to raise polyclonal antibodies in a rabbit. The antiserum was subsequently used in immunohistochemistry assays using the avidin-biotin complex (ABC) method. Cells in the peri-islet region in humans with neoislet formation stained positively for *INGAP* demonstrating that human and hamster *INGAP* share a common epitope between amino acids 104 to 118 in SEQ ID NO:2.

The same synthetic peptide was tested for its ability to stimulate $^{3}$H-thymidine incorporation into rat pancreatic tumor duct cells (ARIP) and hamster insulinoma tumor cells (HIT). 10$\mu$Ci of $^{3}$H-thymidine at 80.4 Ci/mmole concentration was added to approximate $10^{6}$ cells cultured in Ham's F-12K media. After 24 hrs, the cells were harvested and solubilized. Differential precipitation of the nucleic acids with trichloroacetic acid (TCA) was performed according to the procedure modified by Rosenberg et al. and the $^{3}$H-thymidine proportion incorporated was calculated. Addition of the synthetic peptide to ARIP in culture resulted in a 2.4-fold increase in $^{3}$H-thymidine incorporation comparing to the absence of the synthetic peptide in the culture. The synthetic peptide had no effect on the control cell line HIT. This result strongly suggests that *INGAP* plays a role in stimulating islet neogenesis.

References

1. Rosenberg, L., Brown, R.A. and Duguid, W.P. (1982). Surg. Forum 33, 227-230.
2. Rosenberg, L., Brown, R.A. and Duguid, W.P. (1983). J. Surg. Res. 35, 63-72.
3. Rosenberg, L., Duguid, W.P. and Vinik, A.I. (1987). Dig. Dis. Sci. 32, 1185.
4. Clas, D., Rosenberg, L. and Duguid, W.P. (1989). Pancreas 4, 613 (Abstract).
5. Liang, P. and Pardee, B.A. (1992). Science 257, 967-971.
6. Liang, P., Averboukh, L. and Pardee, B.A. (1993). Nucleic Acid Res. 21, 3269-3275.
7. Iovanna, J., Orelle, B., Keim, V. and Dagorn, J.C. (1991). J. Biol. Chem. 266, 24664-24669.
8. Frigerio, J.M., Dusetti, N., Keim, V., Dagorn, J.C. and Iovanna, J. (1993). Biochemistry 32, 9236-9241.
9. Frigerio, J.M., Dusetti, N., Garrido, P., Dagorn, J.C.and Iovanna, J. (1993). Biochim. Biophys. Acta 1216,329-331.
10. Orelle, B., Keim, V., Masciotra, L., Dagorn, J.C. and Iovanna, J. (1992). J. Clin. Invest. 90, 2284-2291.
11. Lasserre, C., Christa, L., Simon, M.T., Vernier, P.and Brechot, C. (1992). Cancer Res. 52, 5089-5095.
12. Drickamer, K. (1988). J. Biol. Chem. 263, 9557-9560.
13. Terazono, K., Yamamoto, H., Takasawa, S., Shiga, K., Yonemura, Y., Tochino, Y. and Okamoto, H. (1988). J. Biol. Chem. 263, 2111-2114.
14. Watanabe, T., Yutaka, Y., Yonekura, H., Suzuki, Y., Miyashita, H., Sugiyama, K., Morizumi, S., Unno, M., Tanaka, O., Kondo, H., Bone, A.J., Takasawa, S. and Okamoto, H. (1994). Proc. Natl. Acad. Sci. USA 91, 3589-3592.
15. Rouquier, S., Giorgi, D., Iovanna, J. and Dagorn, J.C. (1989). Biochem. J. 264, 621-624.
16. Rouquier, S., Verdier, J., Iovanna, J., Dagorn, J.C. and Giorgi, D. (1991) J. Biol. Chem. 266, 786-791.
17. Gross, J., Carlson, R.I., Brauer, A.W., Margolies, M.N., Warshaw, A.L. and Wands, J.R. (1985). J. Clin. Invest. 76, 2115-2126.
18. Iovanna, J., Keim,V., Michael, R.and Dagorn, J.C. (1991). Am. J. Physiol. 261, G485-G489.
19. Rosenberg, L.and Vinik, A.I. (1989). J. Lab. Clin. Med. 114, 75-83.

## SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANTS: Eastern Virginia Medical School of the Medical College of Hampton Roads
McGill University

(ii) TITLE OF INVENTION: INGAP PROTEIN INVOLVED IN PANCREATIC ISLET NEOGENESIS

(iii) NUMBER OF SEQUENCES: 7

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Banner & Allegretti, Ltd.
(B) STREET: 1001 G Street, N.W.
(C) CITY: Washington
(D) STATE: D.C.
(E) COUNTRY: US
(F) ZIP: 20001-4597

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE: 12-FEB-1996
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Kagan, Sarah A.
(B) REGISTRATION NUMBER: 32,141
(C) REFERENCE/DOCKET NUMBER: 00570.54144

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 202-508-9100
(B) TELEFAX: 202-508-9299

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 747 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Cricetulus

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 20..541

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
CTGCAAGACA GGTACCATG ATG CTT CCC ATG ACC CTC TGT AGG ATG TCT TGG          52
                     Met Leu Pro Met Thr Leu Cys Arg Met Ser Trp
                       1               5                  10

ATG CTG CTT TCC TGC CTG ATG TTC CTT TCT TGG GTG GAA GGT GAA GAA          100
Met Leu Leu Ser Cys Leu Met Phe Leu Ser Trp Val Glu Gly Glu Glu
            15                  20                  25

TCT CAA AAG AAA CTG CCT TCT TCA CGT ATA ACC TGT CCT CAA GGC TCT          148
Ser Gln Lys Lys Leu Pro Ser Ser Arg Ile Thr Cys Pro Gln Gly Ser
        30                  35                  40

GTA GCC TAT GGG TCC TAT TGC TAT TCA CTG ATT TTG ATA CCA CAG ACC          196
Val Ala Tyr Gly Ser Tyr Cys Tyr Ser Leu Ile Leu Ile Pro Gln Thr
    45                  50                  55

TGG TCT AAT GCA GAA CTA TCC TGC CAG ATG CAT TTC TCA GGA CAC CTG          244
Trp Ser Asn Ala Glu Leu Ser Cys Gln Met His Phe Ser Gly His Leu
60                  65                  70                  75

GCA TTT CTT CTC AGT ACT GGT GAA ATT ACC TTC GTG TCC TCC CTT GTG          292
Ala Phe Leu Leu Ser Thr Gly Glu Ile Thr Phe Val Ser Ser Leu Val
                80                  85                  90

AAG AAC AGT TTG ACG GCC TAC CAG TAC ATC TGG ATT GGA CTC CAT GAT          340
Lys Asn Ser Leu Thr Ala Tyr Gln Tyr Ile Trp Ile Gly Leu His Asp
            95                  100                 105

CCC TCA CAT GGT ACA CTA CCC AAC GGA AGT GGA TGG AAG TGG AGC AGT          388
Pro Ser His Gly Thr Leu Pro Asn Gly Ser Gly Trp Lys Trp Ser Ser
        110                 115                 120

TCC AAT GTG CTG ACC TTC TAT AAC TGG GAG AGG AAC CCC TCT ATT GCT          436
Ser Asn Val Leu Thr Phe Tyr Asn Trp Glu Arg Asn Pro Ser Ile Ala
    125                 130                 135

GCT GAC CGT GGT TAT TGT GCA GTT TTG TCT CAG AAA TCA GGT TTT CAG          484
Ala Asp Arg Gly Tyr Cys Ala Val Leu Ser Gln Lys Ser Gly Phe Gln
140                 145                 150                 155

AAG TGG AGA GAT TTT AAT TGT GAA AAT GAG CTT CCC TAT ATC TGC AAA          532
Lys Trp Arg Asp Phe Asn Cys Glu Asn Glu Leu Pro Tyr Ile Cys Lys
                160                 165                 170

TTC AAG GTC TAGGGCAGTT CTAATTTCAA CAGCTTGAAA ATATTATGAA                  581
Phe Lys Val

GCTCACATGG ACAAGGAAGC AAGTATGAGG ATTCACTCAG GAAGAGCAAG CTCTGCCTAC      641

ACACCCACAC CAATTCCCTT ATATCATCTC TGCTGTTTTT CTATCAGTAT ATTCTGTGGT      701

GGCTGTAACC TAAAGGCTCA GAGAACAAAA ATAAAATGTC ATCAAC                      747
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 174 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Leu Pro Met Thr Leu Cys Arg Met Ser Trp Met Leu Leu Ser Cys
1 5 10 15

Leu Met Phe Leu Ser Trp Val Glu Gly Glu Glu Ser Gln Lys Lys Leu
20 25 30

Pro Ser Ser Arg Ile Thr Cys Pro Gln Gly Ser Val Ala Tyr Gly Ser
35 40 45

Tyr Cys Tyr Ser Leu Ile Leu Ile Pro Gln Thr Trp Ser Asn Ala Glu
50 55 60

Leu Ser Cys Gln Met His Phe Ser Gly His Leu Ala Phe Leu Leu Ser
65 70 75 80

Thr Gly Glu Ile Thr Phe Val Ser Ser Leu Val Lys Asn Ser Leu Thr
85 90 95

Ala Tyr Gln Tyr Ile Trp Ile Gly Leu His Asp Pro Ser His Gly Thr
100 105 110

Leu Pro Asn Gly Ser Gly Trp Lys Trp Ser Ser Ser Asn Val Leu Thr
115 120 125

Phe Tyr Asn Trp Glu Arg Asn Pro Ser Ile Ala Ala Asp Arg Gly Tyr
130 135 140

Cys Ala Val Leu Ser Gln Lys Ser Gly Phe Gln Lys Trp Arg Asp Phe
145 150 155 160

Asn Cys Glu Asn Glu Leu Pro Tyr Ile Cys Lys Phe Lys Val
165 170

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 175 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Rattus rattus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Leu His Arg Leu Ala Phe Pro Val Met Ser Trp Met Leu Leu Ser
1 5 10 15

Cys Leu Met Leu Leu Ser Gln Val Gln Gly Glu Asp Ser Pro Lys Lys
20 25 30

Ile Pro Ser Ala Arg Ile Ser Cys Pro Lys Gly Ser Gln Ala Tyr Gly
35 40 45

Ser Tyr Cys Tyr Ala Leu Phe Gln Ile Pro Gln Thr Trp Phe Asp Ala
50 55 60

Glu Leu Ala Cys Gln Lys Arg Pro Glu Gly His Leu Val Ser Val Leu
65 70 75 80

Asn Val Ala Glu Ala Ser Phe Leu Ala Ser Met Val Lys Asn Thr Gly
85 90 95

Asn Ser Tyr Gln Tyr Ile Trp Ile Gly Leu His Asp Pro Thr Leu Gly
100 105 110

Gly Glu Pro Asn Gly Gly Gly Trp Glu Trp Ser Asn Asn Asp Ile Met
115 120 125

Asn Tyr Val Asn Trp Glu Arg Asn Pro Ser Thr Ala Leu Asp Arg Gly
130 135 140

Phe Cys Gly Ser Leu Ser Arg Ser Ser Gly Phe Leu Arg Trp Arg Asp
145 150 155 160

Thr Thr Cys Glu Val Lys Leu Pro Tyr Val Cys Lys Phe Thr Gly
165 170 175

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 175 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Leu Pro Pro Met Ala Leu Pro Ser Val Ser Trp Met Leu Leu Ser
1 5 10 15

Cys Leu Met Leu Leu Ser Gln Val Gln Gly Glu Glu Pro Gln Arg Glu
20 25 30

Leu Pro Ser Ala Arg Ile Arg Cys Pro Lys Gly Ser Lys Ala Tyr Gly
35 40 45

Ser His Cys Tyr Ala Leu Phe Leu Ser Pro Lys Ser Trp Thr Asp Ala
50 55 60

Asp Leu Ala Cys Gln Lys Arg Pro Ser Gly Asn Leu Val Ser Val Leu
65 70 75 80

Ser Gly Ala Glu Gly Ser Phe Val Ser Ser Leu Val Lys Ser Ile Gly
85 90 95

Asn Ser Tyr Ser Tyr Val Trp Ile Gly Leu His Asp Pro Thr Gln Gly
100 105 110

Thr Glu Pro Asn Gly Glu Gly Trp Glu Trp Ser Ser Ser Asp Val Met
115 120 125

Asn Tyr Phe Ala Trp Glu Arg Asn Pro Ser Thr Ile Ser Ser Pro Gly
130 135 140

His Cys Ala Ser Leu Ser Arg Ser Thr Ala Phe Leu Arg Trp Lys Asp
145 150 155 160

Tyr Asn Cys Asn Val Arg Leu Pro Tyr Val Cys Lys Phe Thr Asp
165 170 175

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 174 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Rattus rattus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met Leu Pro Arg Val Ala Leu Thr Thr Met Ser Trp Met Leu Leu Ser
1 5 10 15

Ser Leu Met Leu Leu Ser Gln Val Gln Gly Glu Asp Ala Lys Glu Asp
20 25 30

Val Pro Thr Ser Arg Ile Ser Cys Pro Lys Gly Ser Arg Ala Tyr Gly
35 40 45

Ser Tyr Cys Tyr Ala Leu Phe Ser Val Ser Lys Ser Trp Phe Asp Ala
50 55 60

Asp Leu Ala Cys Gln Lys Arg Pro Ser Gly His Leu Val Ser Val Leu
65 70 75 80

Ser Gly Ser Glu Ala Ser Phe Val Ser Ser Leu Ile Lys Ser Ser Gly
85 90 95

Asn Ser Gly Gln Asn Val Trp Ile Gly Leu His Asp Pro Thr Leu Gly
100 105 110

Gln Glu Pro Asn Arg Gly Gly Trp Glu Trp Ser Asn Ala Asp Val Met
115 120 125

Asn Tyr Phe Asn Trp Glu Thr Asn Pro Ser Ser Val Ser Gly Ser His
130 135 140

Cys Gly Thr Leu Thr Arg Ala Ser Gly Phe Leu Arg Trp Arg Glu Asn
145 150 155 160

Asn Cys Ile Ser Glu Leu Pro Tyr Val Cys Lys Phe Lys Ala
165 170

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 174 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Rattus rattus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Leu Pro Arg Leu Ser Phe Asn Asn Val Ser Trp Thr Leu Leu Tyr
1 5 10 15

Tyr Leu Phe Ile Phe Gln Val Arg Gly Glu Asp Ser Gln Lys Ala Val
20 25 30

Pro Ser Thr Arg Thr Ser Cys Pro Met Gly Ser Lys Ala Tyr Arg Ser
35 40 45

Tyr Cys Tyr Thr Leu Val Thr Thr Leu Lys Ser Trp Phe Gln Ala Asp
50 55 60

Leu Ala Cys Gln Lys Arg Pro Ser Gly His Leu Val Ser Ile Leu Ser
65 70 75 80

Gly Gly Glu Ala Ser Phe Val Ser Ser Leu Val Thr Gly Arg Val Asn
85 90 95

Asn Asn Gln Asp Ile Trp Ile Trp Leu His Asp Pro Thr Met Gly Gln
100 105 110

Gln Pro Asn Gly Gly Gly Trp Glu Trp Ser Asn Ser Asp Val Leu Asn
115 120 125

Tyr Leu Asn Trp Asp Gly Asp Pro Ser Ser Thr Val Asn Arg Gly Asn
130 135 140

Cys Gly Ser Leu Thr Ala Thr Ser Glu Phe Leu Lys Trp Gly Asp His
145 150 155 160

His Cys Asp Val Glu Leu Pro Phe Val Cys Lys Phe Lys Gln
165 170

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 165 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Rattus rattus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met Thr Arg Asn Lys Tyr Phe Ile Leu Leu Ser Cys Leu Met Val Leu
1 5 10 15

Ser Pro Ser Gln Gly Gln Glu Ala Glu Glu Asp Leu Pro Ser Ala Arg
20 25 30

Ile Thr Cys Pro Glu Gly Ser Asn Ala Tyr Ser Ser Tyr Cys Tyr Tyr
35 40 45

Phe Met Glu Asp His Leu Ser Trp Ala Glu Ala Asp Leu Phe Cys Gln
50 55 60

Asn Met Asn Ser Gly Tyr Leu Val Ser Val Leu Ser Gln Ala Glu Gly
65 70 75 80

Asn Phe Leu Ala Ser Leu Ile Lys Glu Ser Gly Thr Thr Ala Ala Asn
85 90 95

Val Trp Ile Gly Leu His Asp Pro Lys Asn Asn Arg Arg Trp His Trp
100 105 110

Ser Ser Gly Ser Leu Phe Leu Tyr Lys Ser Trp Asp Thr Gly Tyr Pro
115 120 125

Asn Asn Ser Asn Arg Gly Tyr Cys Val Ser Val Thr Ser Asn Ser Gly
130 135 140

Tyr Lys Lys Trp Arg Asp Asn Ser Cys Asp Ala Gln Leu Ser Phe Val
145 150 155 160

Cys Lys Phe Lys Ala
165

**CLAIMS**

1. A preparation of a mammalian INGAP protein substantially free of other mammalian proteins.

2. The preparation of claim 1 wherein the INGAP protein has the amino acid sequence shown in SEQ ID NO: 2.

3. A preparation of a polypeptide which comprises a sequence of at least 15 consecutive amino acids of a mammalian INGAP protein.

4. The preparation of claim 3 wherein said polypeptide is a fusion of said sequence to a second polypeptide derived from a second protein.

5. The preparation of claim 3 wherein said polypeptide is conjugated to a second polypeptide.

6. The preparation of claim 3 wherein said polypeptide is conjugated to a solid support.

7. The preparation of claim 3 wherein said polypeptide has a biological activity of said mammalian INGAP protein.

8. The preparation of claim 7 wherein said biological activity is the ability to stimulate pancreatic duct cells to grow and proliferate.

9. The preparation of claim 3 wherein said polypeptide comprises amino acids #103 to #122 of the mammalian INGAP protein as shown in SEQ ID NO:2.

10. The preparation of claim 3 wherein said polypeptide comprises at least 130 consecutive amino acids of said mammalian INGAP protein.

11. An isolated DNA molecule encoding a mammalian INGAP protein.

12. The DNA molecule of claim 11 wherein the INGAP protein has the amino acid sequence shown in SEQ ID NO: 2.

13. The DNA molecule of claim 11 wherein the INGAP protein has the nucleotide sequence shown in SEQ ID NO: 1.

14. A vector comprising the DNA of claim 11.

15. The vector of claim 14 further comprising expression control sequences, whereby said DNA is expressed in a host cell.

16. The vector of claim 15 which comprises a EBNA His plasmid.

17. A host cell transformed with the DNA of claim 11.

18. A host cell transformed with the vector of claim 14.

19. The host cell of claim 17 which is a cos7, African Green Monkey kidney cell.

20. A preparation of a mammalian INGAP protein made by the process of:

inducing mammalian pancreatic cells to express INGAP protein by cellophane-wrapping; and

purifying said INGAP protein from said induced mammalian pancreatic cells.

21. A nucleotide probe comprising at least 20 contiguous nucleotides of a mammalian *INGAP* gene.

22. The nucleotide probe of claim 21 wherein the mammalian *INGAP* gene has the sequence shown in SEQ ID NO: 1.

23. The nucleotide probe of claim 21 wherein said probe is labeled with a detectable moiety.

24. A DNA molecule comprising at least 20 contiguous nucleotides of a mammalian *INGAP* gene.

25. The DNA molecule of claim 24 wherein the mammalian *INGAP* gene has the sequence shown in SEQ ID NO: 1.

26. The DNA molecule of claim 24 wherein said molecule is labeled with a detectable moiety.

27. A preparation of an INGAP protein of a mammal substantially purified from other proteins of the mammal wherein said INGAP protein is inducible upon cellophane-wrapping of pancreas of the mammal.

28. A method of isolating an *INGAP* gene from a mammal, comprising:

hybridizing one or more oligonucleotides comprising at least 10 contiguous nucleotides of the sequence shown in SEQ ID NO: 1 to genomic DNA or cDNA of said mammal;

identifying DNA molecules from said genomic DNA or cDNA which hybridize to said one or more oligonucleotides.

29. The method of claim 28 wherein two oligonucleotides are hybridized to said genomic DNA or cDNA of said mammal and said oligonucleotides are used as primers in a polymerase chain reaction (PCR) to synthesize *INGAP* nucleotides from the mammal.

30. The method of claim 28 wherein said one or more oligonucleotides are labelled.

31. The method of claim 28 wherein said genomic DNA or cDNA of said mammal used in said step of hybridizing is in the form of a library of molecular clones.

32. An isolated cDNA molecule obtained by the process of:

hybridizing one or more oligonucleotides comprising at least 10 contiguous nucleotides of the sequence shown in SEQ ID NO: 1 to genomic DNA or cDNA of said mammal;

identifying DNA molecules from said genomic DNA or cDNA which hybridize to said one or more oligonucleotides.

33. An antibody preparation which is specifically immunoreactive with a mammalian INGAP protein.

34. The antibody preparation of claim 33 wherein said mammalian INGAP protein has an amino acid sequence as shown in SEQ ID NO: 2.

35. The antibody preparation of claim 33 which is polyclonal.

36. The antibody preparation of claim 33 which is monoclonal.

37. The antibody preparation of claim 33 comprising antibodies which are bound to a solid phase.

38. A hybridoma which produces antibodies which are specifically immunoreactive with a mammalian INGAP protein.

39. A method of producing a mammalian INGAP protein, comprising the steps of:

providing a host cell according to claim 17;

culturing the host cell in a nutrient medium so that the INGAP protein is expressed; and

harvesting the INGAP protein from the host cells or the nutrient medium.

40. A method of producing a mammalian INGAP protein, comprising the steps of:

providing a host cell comprising the DNA molecule of claim 11;

culturing the host cell in a nutrient medium so that the mammalian INGAP protein is expressed; and

harvesting the mammalian INGAP protein from the host cells or the nutrient medium.

41. A method of treating diabetic mammals, comprising:

administering to a diabetic mammal a therapeutically effective amount of an INGAP protein to stimulate growth of islet cells.

42. The method of claim 41 wherein said mammal has insulin-dependent diabetes mellitus.

43. The method of claim 41 wherein said mammal has non-insulin-dependent diabetes mellitus.

44. A method of growing pancreatic islet cells in culture, comprising:

supplying an INGAP protein to a culture medium for growing pancreatic islet cells; and

growing islet cells in said culture medium comprising INGAP protein.

45. A method of enhancing the life span of pancreatic islet cells encapsulated in a polycarbon shell, comprising:

adding to said encapsulated pancreatic islet cells an INGAP protein in an amount sufficient to enhance the survival rate or survival time of said pancreatic islet cells.

46. A method of enhancing the number of pancreatic islet cells in a mammal, comprising:

administering a DNA molecule which encodes an INGAP protein to a pancreas in a mammal.

47. The method of claim 46 wherein said DNA molecule has the sequence shown in SEQ ID NO:1.

48. The method of claim 46 wherein said INGAP protein has the amino acid sequence shown in SEQ ID NO:2.

49. A method of enhancing the number of pancreatic islet cells in a mammal, comprising:

administering an INGAP protein to a pancreas in a mammal.

50. The method of claim 49 wherein said INGAP protein has the amino acid sequence shown in SEQ ID NO:2.

51. A transgenic mammal which comprises an *INGAP* gene of a second mammal.

52. The transgenic mammal of claim 51 wherein the *INGAP* gene has the sequence shown in SEQ ID NO:1.

53. The transgenic mammal of claim 51 wherein the *INGAP* gene is expressed to a higher level than any endogenous *INGAP* gene of said mammal.

54. A non-human mammal which has been genetically engineered to contain an insertion or deletion mutation of an *INGAP* gene of said mammal.

55. A method of identifying individual mammals at risk for diabetes, comprising:

identifying a mutation in an *INGAP* gene of a sample of an individual mammal, said mutation causing a structural abnormality in an INGAP protein

encoded by said gene or causing a regulatory defect leading to diminished or obliterated expression of said *INGAP* gene.

56. The method of claim 55 wherein said sample is a blood sample.

57. The method of claim 55 wherein said sample is amniotic fluid.

58. The method of claim 55 wherein said sample is chorionic villus.

59. The method of claim 55 wherein said sample is from a blastocyst.

60. The method of claim 55 wherein said sample is pancreatic cells.

61. A method of detecting INGAP protein in a sample from a mammal, comprising:

contacting said sample with an antibody preparation according to claim 33.

62. The method of claim 61 wherein a predetermined amount of a polypeptide comprising at least 15 consecutive amino acids of a mammalian INGAP protein is also contacted with said sample.

63. The method of claim 62 wherein said polypeptide is detectably labeled.

64. The method of claim 61 wherein said antibody preparation comprises antibodies which are bound to a solid support.

65. The method of claim 62 wherein said antibody preparation comprises antibodies which are bound to a solid support.

66. The method of claim 65 further comprising the step of:

detecting labeled polypeptide which is not bound to the solid support.

67. A method of treating isolated islet cells of a mammal to avoid apoptosis of said cells, comprising:

contacting isolated islet cells of a mammal with a preparation of a mammalian INGAP protein, substantially purified from other mammalian proteins, in an amount sufficient to increase the survival rate of said isolated islet cells.

68. A method of treating a mammal receiving a transplant of islet cells, comprising:

administering a preparation of a mammalian INGAP protein to a mammal receiving a transplant of islet cells, wherein said step of administering is performed before, during, or after said transplant.

69. The method of claim 68 wherein said step of administering is performed intravenously.

70. The method of claim 68 wherein said step of administering is performed by local perfusion to the site of said transplant.

71. The method of claim 68 wherein said step of administering is via the portal vein.

72. The method of claim 71 wherein islet cells are concomitantly transplanted via the portal vein.

73. A method of inducing differentiation of $\beta$ cell progenitors, comprising:

contacting a culture of pancreatic duct cells comprising $\beta$ cell progenitors with a preparation of a mammalian INGAP protein substantially free of other mammalian proteins, to induce differentiation of said $\beta$ cell progenitors.

74. A method of treating a mammal with pancreatic endocrine failure, comprising:

contacting a preparation of pancreatic duct cells comprising $\beta$ cell progenitors isolated from a mammal afflicted with pancreatic endocrine failure with a preparation of a mammalian INGAP protein substantially free of other mammalian proteins to induce differentiation of said $\beta$ cell progenitors; and

autologously transplanting said treated pancreatic duct cells into said mammal.

75. An antisense construct of a mammalian *INGAP* gene comprising:

a promoter, a terminator, and a nucleotide sequence consisting of a mammalian *INGAP* gene, said nucleotide sequence being between said promoter

and said terminator, said nucleotide sequence being inverted with respect to said promoter, whereby upon expression from said promoter an mRNA complementary to native mammalian *INGAP* mRNA is produced.

76. A method of treating nesidioblastosis comprising:

administering to a mammal with nesidioblastosis an antisense construct according to claim 75, whereby overgrowth of $\beta$ cells of said mammal is inhibited.

77. A kit for detecting a mammalian INGAP protein in a sample from a mammal, comprising:

an antibody preparation which is specifically immunoreactive with a mammalian INGAP protein;

a polypeptide which comprises a sequence of at least 15 consecutive amino acids of a mammalian INGAP protein.

78. The kit of claim 77 wherein said polypeptide is detectably labeled.

79. The kit of claim 77 wherein said antibody preparation comprises antibodies which are bound to a solid support.

80. A pharmaceutical composition for treatment of pancreatic insufficiency, comprising:

a mammalian INGAP protein in a pharmaceutically acceptable diluent or carrier.

81. The pharmaceutical composition of claim 80 wherein the INGAP protein has the amino acid sequence shown in SEQ ID NO: 2.

82. A pharmaceutical composition comprising:

a preparation of a polypeptide which comprises a sequence of at least 15 consecutive amino acids of a mammalian INGAP protein and a pharmaceutically acceptable diluent or carrier.

83. The pharmaceutical composition of claim 82 wherein said polypeptide is a fusion of said sequence to a second polypeptide derived from a second protein.

84. The pharmaceutical composition of claim 82 wherein said polypeptide is conjugated to a second polypeptide.

85. The pharmaceutical composition of claim 82 wherein said polypeptide has a biological activity of said mammalian INGAP protein.

86. The pharmaceutical composition of claim 85 wherein said biological activity is the ability to stimulate pancreatic duct cells to grow and proliferate.

87. The pharmaceutical composition of claim 82 wherein said polypeptide comprises amino acids #103 to #122 of the mammalian INGAP protein as shown in SEQ ID NO:2.

88. The pharmaceutical composition of claim 82 wherein said polypeptide comprises at least 130 consecutive amino acids of said mammalian INGAP protein.

89. A method of identifying $\beta$ cell progenitors, comprising:

contacting a population of pancreatic duct cells with a preparation of a mammalian INGAP protein; and

detecting cells from among said population to which said INGAP specifically binds.

90. The method of claim 89 wherein said INGAP protein is detectably labeled.

91. The method of claim 89 wherein said INGAP protein is immobilized on a solid phase.

92. The preparation of claim 1 wherein the INGAP protein is from human and comprises amino acid sequences 1 to 83 and 124 to 174 as shown in SEQ ID NO:2.

93. The preparation of claim 1 wherein the INGAP protein is from human and comprises in a N-terminal to C-terminal orientation: amino acids 1 to 83 in SEQ ID NO:2, 40 amino acids, and amino acids 124 to 174 in SEQ ID NO:2.

94. The DNA molecule of claim 11 wherein the INGAP protein is from human.

95. The DNA molecule of claim 94 wherein said INGAP protein comprises amino acid sequences 1 to 83 and 124 to 174 in SEQ ID NO:2.

96. The DNA molecule of claim 94 wherein said INGAP protein comprises in an N-terminal to C-terminal orientation amino acids: 1 to 83 in SEQ ID NO:2, 40 amino acids, and amino acids 124 to 174 in SEQ ID NO:2.

97. The DNA molecule of claim 24 which encodes an amino acid sequence selected from those of amino acids 1 to 83 and 124 to 174 in SEQ ID NO:2.

98. The DNA molecule of claim 11 which comprises nucleotides 4 to 268 and 389 to 629 of SEQ ID NO:1.

# FIG. 1A

```
CTGCAAGACA GGTACCATG ATG CTT CCC ATG ACC CTC TGT AGG ATG TCT TGG          52
                     Met Leu Pro Met Thr Leu Cys Arg Met Ser Trp
                       1               5                  10

ATG CTG CTT TCC TGC CTG ATG TTC CTT TCT TGG GTG GAA GGT GAA GAA          100
Met Leu Leu Ser Cys Leu Met Phe Leu Ser Trp Val Glu Gly Glu Glu
            15                  20                  25

TCT CAA AAG AAA CTG CCT TCT TCA CGT ATA ACC TGT CCT CAA GGC TCT          148
Ser Gln Lys Lys Leu Pro Ser Ser Arg Ile Thr Cys Pro Gln Gly Ser
        30                  35                  40

GTA GCC TAT GGG TCC TAT TGC TAT TCA CTG ATT TTG ATA CCA CAG ACC          196
Val Ala Tyr Gly Ser Tyr Cys Tyr Ser Leu Ile Leu Ile Pro Gln Thr
    45                  50                  55

TGG TCT AAT GCA GAA CTA TCC TGC CAG ATG CAT TTC TCA GGA CAC CTG          244
Trp Ser Asn Ala Glu Leu Ser Cys Gln Met His Phe Ser Gly His Leu
60                  65                  70                  75

GCA TTT CTT CTC AGT ACT GGT GAA ATT ACC TTC GTG TCC TCC CTT GTG          292
Ala Phe Leu Leu Ser Thr Gly Glu Ile Thr Phe Val Ser Ser Leu Val
                80                  85                  90

AAG AAC AGT TTG ACG GCC TAC CAG TAC ATC TGG ATT GGA CTC CAT GAT          340
Lys Asn Ser Leu Thr Ala Tyr Gln Tyr Ile Trp Ile Gly Leu His Asp
            95                  100                 105
```

# FIG. 1B

```
CCC TCA CAT GGT ACA CTA CCC AAC GGA AGT GGA TGG AAG TGG AGC AGT        388
Pro Ser His Gly Thr Leu Pro Asn Gly Ser Gly Trp Lys Trp Ser Ser
        110                 115                 120

TCC AAT GTG CTG ACC TTC TAT AAC TGG GAG AGG AAC CCC TCT ATT GCT        436
Ser Asn Val Leu Thr Phe Tyr Asn Trp Glu Arg Asn Pro Ser Ile Ala
    125                 130                 135

GCT GAC CGT GGT TAT TGT GCA GTT TTG TCT CAG AAA TCA GGT TTT CAG        484
Ala Asp Arg Gly Tyr Cys Ala Val Leu Ser Gln Lys Ser Gly Phe Gln
140                 145                 150                 155

AAG TGG AGA GAT TTT AAT TGT GAA AAT GAG CTT CCC TAT ATC TGC AAA        532
Lys Trp Arg Asp Phe Asn Cys Glu Asn Glu Leu Pro Tyr Ile Cys Lys
                160                 165                 170

TTC AAG GTC TAGGGCAGTT CTAATTTCAA CAGCTTGAAA ATATTATGAA               581
Phe Lys Val

GCTCACATGG ACAAGGAAGC AAGTATGAGG ATTCACTCAG GAAGAGCAAG CTCTGCCTAC     641

ACACCCACAC CAATTCCCTT ATATCATCTC TGCTGTTTTT CTATCAGTAT ATTCTGTGGT     701

GGCTGTAACC TAAAGGCTCA GAGAACAAAA ATAAAATGTC ATCAAC                    747
```

# FIG. 2

```
INGAP          MLPMTLC-RMSWMLLSCLMFLSWVEGEESQKKLPSS   35
PAP-I          MLHRLAFPVMSWMLLSCLMLLSQVQGEDSPKKIPSA   36
PAP-H/HIP      MLPPMALPSVSWMLLSCLMLLSQVQGEEPQRELPSA   36
PAP-III        MLPRVALTTMSWMLLSSLMLLSQVQGEDAKEDVPTS   36
PAP-II         MLPRLSFNNVSWTLLYYLFIF-QVRGEDSQKAVPST   35
REG/LITH       ----MT-RNKYFILLSCLMVLSPSQGQEAEEDLPSA   31
"DRICKAMER"

                   *          *              *
INGAP          RITCPQGSVAYGSYCYSLILIPQTWSNAELSCQMHF   71
PAP-I          RISCPKGSQAYGSYCYALFQIPQTWFDAELACQKRP   72
PAP-H/HIP      RIRCPKGSKAYGSHCYALFLSPKSWTDADLACQKRP   72
PAP-III        RISCPKGSRAYGSYCYALFSVSKSWFDADLACQKRP   72
PAP-II         RTSCPMGSKAYRSYCYTLVTTLKSWFQADLACQKRP   71
REG/LITH       RITCPEGSNAYSSYCYYFMEDHLSWAEADLFCQNMN   67
"DRICKAMER"          G                        C

INGAP          SGHLAFLLSTGEITFVSSLVKNSLTAYQYIWIGLHD   107
PAP-I          EGHLVSVLNVAEASFLASMVKNTGNSYQYIWIGLHD   108
PAP-H/HIP      SGNLVSVLSGAEGSFVSSLVKSIGNSYSYVWIGLHD   108
PAP-III        SGHLVSVLSGSEASFVSSLIKSSGNSGQNVWIGLHD   108
PAP-II         SGHLVSILSGGEASFVSSLVTGRVNNNQDIWIWLHD   107
REG/LITH       SGYLVSVLSQAEGNFLASLIKESGTTAANVWIGLHD   103
"DRICKAMER"                                 G TD

INGAP          PSHGTLPNGSGWKWSSSNVLTFYNWERNPSIAADRG   143
PAP-I          PTLGGEPNGGGWEWSNNDIMNYVNWERNPSTALDRG   144
PAP-H/HIP      PTQGTEPNGEGWEWSSSDVMNYFAWERNPSTISSPG   144
PAP-III        PTLGQEPNRGGWEWSNADVMNYFNWETNPSSVSGS-   143
PAP-II         PTMGQQPNGGGWEWSNSDVLNYLNWDGDPSSTVNRG   143
REG/LITH       P-----KNNRRWHWSSGSLFLYKSWDTGYPNNSNRG   134
"DRICKAMER"                  T        W    P     G

                   *          *          *
INGAP          YCAVLSQKSGFQKWRDFNCENELPYICKFKV   175
PAP-I          FCGSLSRSSGFLRWRDTTCEVKLPYVCKFTG   176
PAP-H/HIP      HCASLSRSTAFLRWKDYNCNVRLPYVCKFTD   176
PAP-III        HCGTLTRASGFLRWRENNCISELPYVCKFKA   175
PAP-II         NCGSLTATSEFLKWGDHHCDVELPFVCKFKQ   175
REG/LITH       YCVSVTSNSGYKKWRDNSCDAQLSFVCKFKA   165
"DRICKAMER"    EC        G   WND  C        CE
```

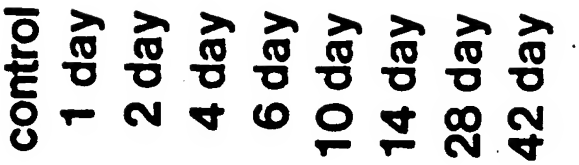


FIG. 3A


FIG. 3B


FIG. 3C

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/01528

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(6) :Please See Extra Sheet.
US CL :Please See Extra Sheet.
According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 530/350, 387.1; 536/23.5, 24.5; 435/240.2, 320.1, 172.3, 6, 7.1, 91.2; 514/2, 44; 800/2; 424/93.7

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, DIALOG
search terms: pancreas, INGAP, gene, DNA, cloning, neogenesis, REG, diabetes, inventor's name

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | ROSENBERG et al. Reversal of diabetes by the induction of Islet cell neogenesis. Transplantation Proceedings. June 1992, Vol. 24, No. 3, pages 1027-1028, see the entire document. | 1-98 |
| Y | LIANG et al. Distribution and cloning of eukaryotic mRNAs by means of differential display: refinements and optimization. Nucleic Acids Research. 1993, Vol. 21, No. 14, pages 3269-3275, see the entire document. | 1-98 |
| Y | US 4,965,188 A (K.B. MULLIS ET AL.) 23 October 1990, entire document. | 28-32, 55-60 |

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
27 MAY 1996

Date of mailing of the international search report
07 JUN 1996

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
JASEMINE C. CHAMBERS

Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/01528

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | WATANABE et al. Pancreatic beta-cell replication and amelioration of surgical diabetes by Reg protein. Proc. Natl. Acad. Sci. USA. April 1994, Vol. 91, pages 3589-3592, see the entire document. | 41-43, 49, 50, 67-72, 74, 80-88 |
| Y | MILLER. Human gene therapy comes of age. Nature. 11 June 1992, Vol. 357, pages 455-460, see the entire document. | 46-48 |
| Y | BRADLEY et al. Bio/Technology. Modifying the Mouse: Design and Desire, May 1992, Vol. 10, pages 534-539, see the entire document. | 51-54 |
| Y | STEIN et al. Antisense oligonucleotides as therapeutic agents - is the bullet really magical? Science, 20 August 1993, Vol. 261, pages 1004-1012, see the entire document. | 75, 76 |

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/01528

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (6):

C07K 14/00, 16/00; C12N 5/00, 15/00; C07H 21/00; A61K 31/00, 38/00, 48/00; C12Q 1/68; C12P 19/34; G01N 33/00

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

530/350, 387.1; 536/23.5, 24.5; 435/240.2, 320.1, 172.3, 6, 7.1, 91.2; 514/2, 44; 800/2; 424/93.7